

## Інтервальне нечітке моделювання складних систем в умовах недовизначеності вхідних даних

Н. Р. Кондратенко, О. О. Снігур

*Запропоновано систему моделей на основі інтервальної нечіткої логічної системи класифікації, що дозволяє отримати вихід за умов відсутності частини вхідних значень. Система будується виходячи з експериментальних даних, допускає залучення одного або кількох експертів, а також інтеграцію сторонніх підмоделей на основі інших методів та технологій прийняття рішень*

*Ключові слова: нечітка логічна система, інтервальні нечіткі множини, недовизначеність, кластерний аналіз*

### 1. Вступ

Недовизначеність вхідних даних – проблема, характерна для моделювання складних природних систем. До таких належать екологічні, соціальні, економічні, технічні системи різної природи, для математичного опису функціонування яких складно побудувати один аналітичний закон, і про характер роботи системи можна судити лише за вибіркою експериментальних даних. Виділяють низку т. зв. НЕ-факторів [1], що мають визначальний вплив на якість вихідної експериментальної вибірки, як-то неточність вимірювань, відсутність можливостей для безпосереднього спостереження об'єкта, неповнота та неоднозначність знань про предметну галузь і розв'язувану задачу, вплив неврахованих (прихованих) параметрів, брак експертних знань про предметну галузь або неможливість їх формалізувати, а також невизначеність, пов'язана з розмірністю простору вхідних ознак (надлишковість та зашумленість) [1, 2].

Всі ці фактори тією чи іншою мірою властиві природним системам та процесам. Особливості моделювання систем цього класу далі буде показано на прикладі оцінювання стану артезіанської свердловини в довільний момент часу в межах від початку гідрогеологічної розвідки до її повного завершення. Дана задача характеризується важкодоступністю експериментальних даних, оскільки отримання вхідних даних, необхідних для роботи будь-якої моделі, вимагає виконання значного обсягу робіт. Невипадково тривалість геологорозвідувальних робіт, що передують введенню свердловини в експлуатацію, складає від 6 місяців до кількох років [3].

У ході геологорозвідувальних робіт (ГРР) можна умовно виділити 3 етапи: попередня розвідка, детальна розвідка та експлуатаційна розвідка [3, 4]. Особливістю артезіанської свердловини як об'єкта моделювання є недовизначеність, а саме наявність пропусків у векторі вхідних даних, що її характеризує в кожен конкретний момент часу аж до повного завершення всіх трьох етапів дослідження. Це зумовлює не-

обхідність розробки математичних моделей та технологій, здатних отримувати результат на ранніх етапах роботи з системою, коли дослідник не володіє повним вектором даних. Такі моделі дадуть змогу на ранніх етапах оцінити перспективи подальшої роботи з системою та виявити випадки, коли подальша робота пов'язана з певними труднощами. На основі цієї інформації можуть бути проведені додаткові дослідження або прийняте рішення про припинення робіт, що дозволить зберегти матеріальні та людські ресурси.

## **2. Аналіз літературних даних та постановка проблеми**

Статистичні моделі, що з успіхом застосовуються для дослідження складних систем різної природи за умов повної визначеності вхідних даних, на недовизначених даних показують себе непрацездатними. Особливо небезпечне застосування статистичних моделей на невеликих вибірках експериментальних даних, оскільки отримані на них закони розподілу можуть бути нестійкими [5]. Крім того, статистичні моделі та методи не враховують або неповною мірою враховують експертні знання.

Технології штучного інтелекту, такі як класифікація і кластерний аналіз, показують добрі результати на даних високої розмірності [6–9]; вони успішно використовуються для виділення прихованих закономірностей та внутрішніх взаємозв'язків [6, 10–12].

Апарат інтервальних нечітких множин дає змогу будувати математичні моделі складних систем та процесів, здатні оперувати вхідними даними, що містять пропуски [13]. Важливою їх особливістю є відсутність обмежень на складність моделі, оскільки моделюється відгук системи на вектор вхідних даних, без необхідності побудови фізичної моделі внутрішніх процесів, що відбуваються в системі, та причинно-наслідкових зв'язків [6]. На сьогоднішній день такі моделі мають широке застосування в медичній діагностиці [13], розпізнаванні образів [14, 15], для моделювання та класифікації мультимедійного трафіку [16], портфельної оптимізації [17], прогнозування часових послідовностей [18, 19] та ін. Оцінка перспективності артезіанської свердловини – ще одна прикладна задача, для якої можливо адаптувати існуючі моделі прийняття рішень на основі нечітких множин та запропонувати нові.

Нечіткі та нейро-нечіткі методи вже довгий час застосовується для оцінювання якості води: методи і системи підтримки прийняття рішень шляхом побудови нечіткого логічного висновку представлено в роботах [20–22]; в [23] запропоновано метод нечіткого оцінювання якості води для множини точок моніторингу. Методи нечіткої кластеризації також дають позитивні результати в цій задачі [24, 25]. Робота з підземними водами ускладнюється їхньою недоступністю для безпосередніх досліджень. На початкових етапах гідрогеологічної розвідки технічно можливо зібрати лише опосередковані знання. Інформація про стан системи достеменно відома в окремих точках родовища; дані ж про інші ділянки отримують, екстраполюючи фактичні точкові дані на ділянки, про які фактичної інформації немає [3]. Тому сучасні методи та технології оцінювання якості підземних вод [26–28] в ці-

лому суттєво не відрізняються від методів, що застосовуються для досліджень поверхневих вод. Вони також вимагають прямого доступу до водоносних горизонтів, можливості проведення пробних і дослідних відкачок, безперешкодного взяття зразків води, тобто повного завершення ГРР. З усіх факторів, що впливають на якість підземних вод та можливість видобутку питної води особливу увагу приділено антропогенному забрудненню [29, 30] та дослідженню вразливості водоносних горизонтів до шкідливих речовин, присутніх у повітрі, ґрунтах та поверхневих водах [31, 32].

З цього погляду моделі на основі нечітких множин типу 1 мають суттєве обмеження: вони не можуть напряду працювати на неповних / недовизначених вхідних даних [13]. Існуючий механізм логічного висновку не дозволяє визначити вихідне значення при неповному вхідному векторі. Інтервальні нечіткі моделі типу 2 дають змогу враховувати і моделювати різні типи невизначеностей, включаючи, в окремих випадках, недовизначеність, породжену відсутністю частини значень [33]. З огляду на це, для прийняття рішень у таких умовах доцільно застосувати математичний апарат нечітких множин типу 2.

Як правило, для вивчення поведінки складних систем будується єдина модель та єдиний критерій оцінювання невідповідності моделі даним спостережень. Такий підхід працює лише тоді, коли між входами та виходами системи є функціональна залежність, та спостереження проводяться точно. Якщо хоча б одна з цих умов не виконується, рекомендується побудова системи моделей та критеріїв відповідності [34, 35]. Природні системи та процеси належать саме до таких, в силу важкодоступності даних спостережень та неможливості гарантувати точність вимірювання кількісних показників. При цьому, відповідно до [34], чим складніша система та чим менша достовірність і доступність даних спостережень, тим більш відмінними будуть моделі, обрані за різними критеріями. Цей фактор також повною мірою стосується природних систем. Виходячи з цього, пропонується не обмежуватись лише можливостями математичного апарату нечітких множин, а запропонувати підхід на основі системи моделей, що допускає використання інших методів і технологій прийняття рішень.

Поставлену задачу можна сформулювати таким чином. Нехай задано набір експериментальних даних  $(X, Y)$ :

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_m^1 \\ x_1^2 & x_2^2 & & x_m^2 \\ \dots & & & \\ x_1^n & x_2^n & \dots & x_m^n \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix},$$

де  $X^i = \{x_1^i, \dots, x_m^i\}$  – результати дослідження системи  $W^i$  за параметрами  $p_1 \dots p_m$ . Для кожного вектора  $X^i$  відомо значення лінгвістичної змінної  $Y$  – остаточний ви-

сненок експерта для системи  $W^i$  (діагноз, клас якості, справність – залежно від прикладного застосування).

Для довільної системи  $W^z$ , що описується вхідним вектором

$$X^z = \{x_1^z, \dots, x_m^z\}, \quad X^z \notin X; \quad x_{k < m}^z, \dots, x_m^z \in \emptyset,$$

необхідно визначити значення лінгвістичної змінної. З огляду на описані вище проблеми, задача вимагає побудови системи моделей з можливістю побудови інтервального нечіткого логічного висновку, врахування експертних знань та включення в процес прийняття рішення додаткових моделей і методів.

### **3. Мета та задачі дослідження**

Метою дослідження є розширення можливостей існуючих моделей та методів прийняття рішень в умовах недовизначеності вхідних даних. Пропонується побудувати систему моделей, що об'єднує в собі як переваги системи підтримки прийняття рішень на основі інтервального нечіткого логічного висновку, так і технологій інтелектуального аналізу даних.

Для досягнення поставленої мети вирішувались такі задачі:

- розробити модель класифікації даних, пристосовану до роботи з недовизначеним вхідним вектором;
- запропонувати шляхи врахування експертних знань у процесі прийняття рішення;
- запропонувати агрегований критерій, що дає узагальнену інтервальну оцінку значення вихідної змінної на основі множини моделей;
- запропонувати альтернативну модель прийняття рішення на основі технологій інтелектуального аналізу даних.

### **4. Методи та засоби моделювання стану природної системи в умовах недовизначеності вихідних даних**

Загальний вигляд агрегованої моделі прийняття рішень в умовах недовизначеності вхідних даних показано на рис. 1.

Вихідний вектор експериментальних даних, що може містити пропуски, подається на вхід інтервальної нечіткої логічної системи типу 2 без змін, а також пройшовши через процедуру виділення інформативних ознак. Зазначена процедура може виконуватись одним або кількома експертами в предметній галузі. У разі, коли експертів декілька, кожен із них генерує свій набір ознак, та як наслідок окрему модель. Інтервальний вихід нечіткої логічної системи за повним вхідним вектором та виходи результуючих моделей після відкидання неінформативних ознак об'єднуються за агрегуючим критерієм  $Y_A$ .

Пропонована система моделей допускає існування альтернативних моделей на основі інших технологій прийняття рішень або формальних процедур, що вже

існують для розв'язання поставленої задачі. Далі буде показано процедуру прийняття рішень на основі методу кластерного аналізу в ролі такої моделі.

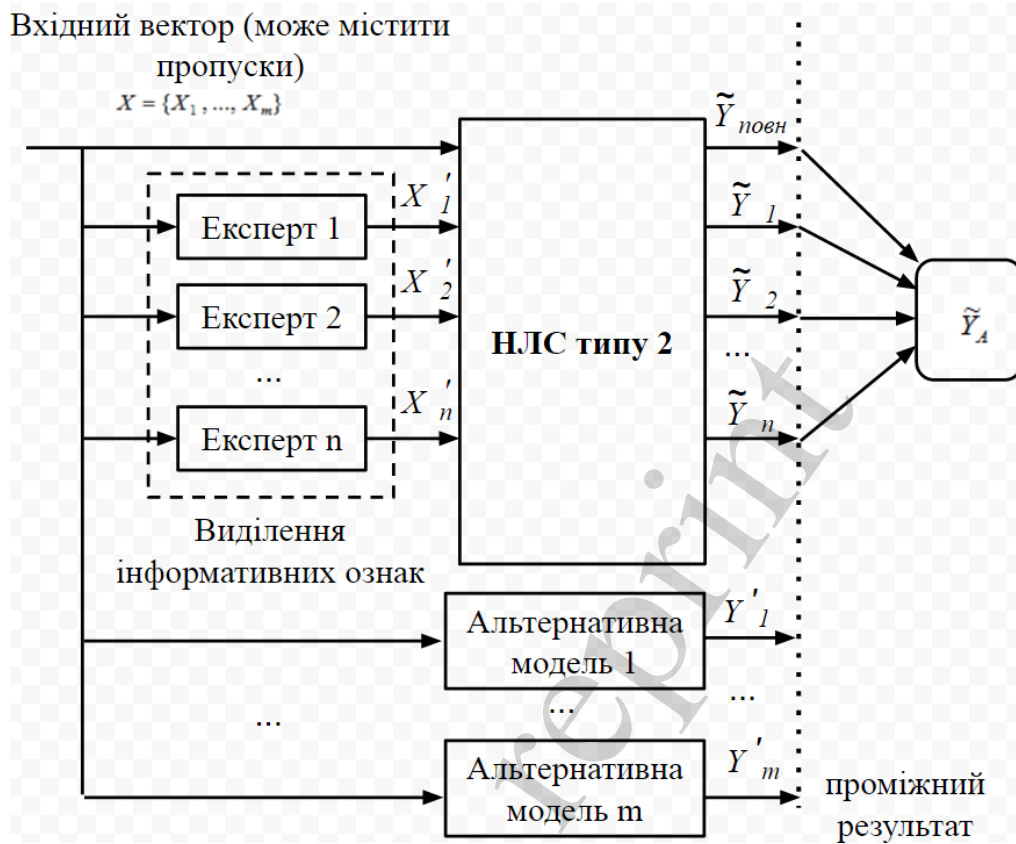


Рис. 1. Агрегована модель прийняття рішень

У випадку, коли інтервальний вихід альтернативної моделі має таку саму розмірність і якісну природу, що й інтервали на виході нечіткої логічної системи, він також враховується при побудові агрегованого інтервалу за критерієм  $Y_A$ . В іншому разі виходи альтернативних моделей подаються користувачу системи окремими інтервалами, незалежно від основного виходу.

#### 4. 1. Прийняття рішень на основі інтервальної нечіткої логічної системи типу 2

Для розв'язання поставленої задачі побудовано нечітку логічну систему-класифікатор. Вхідний вектор являє собою набір значень параметрів системи  $X^i = \{x_1^i, \dots, x_m^i\}$ . База правил формується з відомих пар  $(X_i, Y_i)$ , де вхідному вектору поставлено у відповідність лінгвістичну оцінку значення вихідної змінної  $Y$ , дану експертом. Таким чином, кожен вхідний вектор експериментальних даних генерує одне правило. Антецеденти правил утворюються заміною значення  $x_j^i$  відповідним

йому нечітким термом  $A_{x_j}^i$ , консеквентами є терм лінгвістичної змінної  $y$ , визначений експертом для вектора  $X^i$ :

$$R^i : IF \ x_1 \in A_{x_1}^i \wedge x_2 \in A_{x_2}^i \wedge \dots \wedge x_m \in A_{x_m}^i THEN \ y \in L_y^k \in \{L_1, \dots, L_p\},$$

де  $x_i$  – вхідні змінні,  $y$  – вихідна змінна,  $L_y \in \{L_1, \dots, L_p\}$  – терм-множини вихідної змінної. Терм-множини вхідних та вихідної змінної описуються гаусівськими функ-

ціями належності вигляду  $\mu(x_i^j) = e^{-\left(\frac{x-b_i^j}{c_i^j}\right)^2}$ .

На основі набору значень  $(X, Y)$ , а також бази правил, синтезованої на основі набору експериментальних даних, будується нечітка логічна система типу 1 з чітким виходом  $Y \in [0;10]$ . У разі необхідності для покращення адекватності відображення моделлю навчальних даних проводиться оптимізація параметрів функцій належності. Після цього виконується перетворення результуючих функцій належності типу 1 на інтервальні функції належності типу 2 з невизначеним середнім (рис. 2).

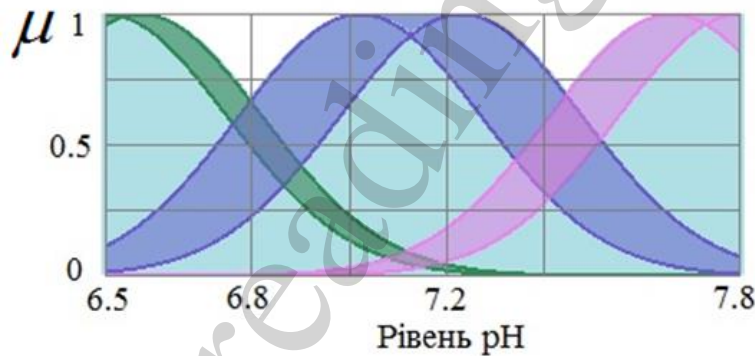


Рис. 2. Приклад інтервальної функції належності з невизначеним середнім

Розширення границь інтервалів функцій належності виконується доти, поки виконується умова:

$$\forall x_i \in X \ F(x_i, P^{(k)}) = F(x_i, P^{(1)}),$$

де  $P^{(1)}$  – початкові параметри функцій належності,

$$P^{(k)} = \{\mu_1 \cdot k, \dots, \mu_p \cdot k\}, \quad k = k \pm 0,001$$

$F(x_i, P^{(k)})$  – вихід системи без дефазифікації – номер терма з максимальним покриттям результуючою функцією належності.

Нечіткий логічний висновок будується за алгоритмом Карніка-Менделя [36]. Інтервальні ступені належності кожного правила розраховуються як мінімум усіх антецедентів:

$$\mu_{R_i} = \left[ \min_j (\underline{\mu}^{(2)A_j^{(i)}}(x_j^*)); \min_j (\bar{\mu}^{(2)A_j^{(i)}}(x_j^*)) \right].$$

Для знаходження лівої та правої границь інтервалу вихідної змінної  $[y_l; y_r]$  на основі розрахованих ступенів належності правил та інтервальних значень консеквентів правил будується вихідна нечітка множина типу 2. Вихідний інтервал значень отримуємо за допомогою процедури пониження типу нечіткої множини. Для правої границі інтервалу:

1. Обчислити

$$f_r^i = \frac{(\underline{\mu}_i + \bar{\mu}_i)}{2}; y_r = \frac{\sum_{i=1}^M f_r^i w_r^i}{\sum_{i=1}^M f_r^i}; y'_r = y_r.$$

2. Знайти  $R$  ( $R=1 \dots M-1$ ) таке, що  $w_r^R \leq y'_r \leq w_r^{R+1}$ .

$$3. y_r = \frac{\sum_{i=1}^R f^i w_r^i + \sum_{i=R+1}^M \bar{f}^i w_r^i}{\sum_{i=1}^R f^i + \sum_{i=R+1}^M \bar{f}^i}; y''_r = y_r.$$

4. Якщо  $y''_r \neq y_r$ , перейти на крок 5, інакше  $y_r = y''_r$  та перейти на крок 6.

5.  $y'_r = y''_r$ , перейти на крок 2.

Для лівої границі інтервалу:

1. Обчислити

$$f_l^i = \frac{(\underline{\mu}_i + \bar{\mu}_i)}{2}; y_l = \frac{\sum_{i=1}^M f_l^i w_l^i}{\sum_{i=1}^M f_l^i}; y'_l = y_l.$$

2. Знайти  $L$  ( $L=1 \dots M-1$ ) таке, що  $w_l^L \leq y'_l \leq w_l^{L+1}$ .

$$3. y_l = \frac{\sum_{i=1}^L \bar{f}^i w_l^i + \sum_{i=L+1}^M \underline{f}^i w_l^i}{\sum_{i=1}^L \bar{f}^i + \sum_{i=L+1}^M \underline{f}^i}; y_l'' = y_l.$$

4. Якщо  $y_l'' \neq y_l$ , перейти на крок 5, інакше  $y_l = y_l''$  та перейти на крок 6.

5.  $y_l' = y_l''$ , повернутися на крок 2.

Ширина отриманого інтервалу  $[y_l'; y_l'']$  характеризує ступінь невизначеності, пов'язаної з прийнятим рішенням.

#### 4. 2. Моделі з пониженням розмірності простору вхідних ознак

Експериментальні дослідження показують, що інтервальний нечіткий класифікатор, описаний вище, не завжди дає бажаний результат. Підмножиною задач моделювання систем складно формалізованої природи є задачі з великою кількістю вхідних параметрів. У таких задачах присутня невизначеність, пов'язана з розмірністю простору вхідних ознак, деякі з яких надлишкові, інші недостатньо інформативні та є джерелом шуму й аномалій в експериментальній вибірці. У випадках, коли робота системи на всіх множині вхідних ознак практично неможлива, пропонується понизити розмірність задачі, відкинувши частину ознак, що не чинять видимого впливу на результат роботи системи. В загальній системі моделей на рис. 1 цю функцію покладено на експертів  $1, \dots, n$ , кожен із яких пропонує свій набір інформативних ознак. Вхідні змінні, відкинуті експертом, виключаються зі вхідного вектора, а також із правил бази знань, в антецедентах яких вони фігурують.

Залучення кількох або навіть одного експерта не завжди видається можливим, тому в ролі експерта можливе використання методу [37], який дозволяє враховувати як теоретичні знання експерта, так і кількісні дані, накопичені на основі спостережень за реальними об'єктами.

#### 4. 3. Прийняття рішень на основі методу нечіткої кластеризації з інтервальними ступенями належності

Окрім механізму прийняття рішень на основі інтервальних нечітких множин, що є основним в описуваній системі моделей, пропонується альтернативна модель на основі модифікованого методу кластеризації РСМ з інтервальним виходом [11]. У термінах кластерного аналізу поставлену задачу можна сформулювати таким чином.

Нехай задано набір експериментальних даних  $X$ :

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_m^1 \\ x_1^2 & x_2^2 & & x_m^2 \\ \dots & & & \\ x_1^n & x_2^n & \dots & x_m^n \end{bmatrix},$$



де  $X^i = \{x_1^i, \dots, x_m^i\}$  – результати дослідження системи  $W^i$  за параметрами  $p_1 \dots p_m$ . В загальному випадку висновок про належність системи  $W^i$  до одного чи кількох класів відповідно до оцінюваного параметру (якості, діагнозу тощо) для кожного з векторів  $X^i$  невідомий, але відомо, що об'єкти  $x^1, \dots, x^n$  утворюють компактні кластери в просторі вхідних ознак  $p_1 \dots p_m$ .

Необхідно розбити множину  $X$  на  $s$  кластерів та визначити ступені належності до кожного з  $s$  кластерів довільної системи  $W^z$ , що описується вхідним вектором  $X^z = \{x_1^z, \dots, x_m^z\}$ ,  $X^z \notin X$ . На множину  $X$  накладається умова репрезентативності її відносно генеральної сукупності векторів ознак, тобто множина  $X$  повинна містити представників усіх  $s$  класів.

Процес прийняття рішення виглядає як процес побудови розбиття на кластери множини, що складається з експериментального набору  $X$  та вектора параметрів досліджуваної системи  $X^z$  за методом [11]. Остаточне рішення приймається за ступенями належності точки, що характеризується вектором  $X^z$ , до кожного з  $s$  утворених кластерів.

#### 4. 4. Агрегований критерій

Від початку до кінця робота описаної системи моделей являє собою таку послідовність дій.

1. Вибрати всі параметри, значення яких відомі на даний момент.
2. Виключити всі інші (невідомі) параметри з антецедент правил.
3. Подати вектор із відомих значень на вхід нечіткої логічної системи типу 2:  
 $X \rightarrow \tilde{Y}_{full}$ .
4. Відкинути (будь-яким доступним шляхом) неінформативні та малоінформативні ознаки та подати результуючий вектор на вхід нечіткої логічної системи типу 2, результатом чого є модель  $X'_i \rightarrow \tilde{Y}_i$ .
5. Повторити крок 4 для всіх доступних способів вибору інформативних ознак.
6. (необов'язковий) Отримати проміжні розв'язки за альтернативними моделями 1..m.
7. Отримати агрегований розв'язок основної моделі за правилом  

$$\tilde{Y}_A = \tilde{Y}_{full} \cap \left( \bigcup_{i=1}^n \tilde{Y}_i \right).$$

### 5. Результати моделювання на прикладі оцінювання стану артезіанської свердловини

#### 5. 1. Інтервальна нечітка кластеризація в задачі оцінювання стану артезіанської свердловини

Застосуємо метод інтервальної кластеризації в технологічній задачі експертного оцінювання стану артезіанської свердловини. Об'єктом кластеризації вважа-

ємо набір значень параметрів артезіанської свердловини  $X^i = \{x_1^i, \dots, x_m^i\}$ , включаючи такі, що описують особливості геологічної будови, тектонічні, кліматичні та гідрогеологічні умови, результати обстеження експлуатаційних свердловин у районі, суміжному родовищу, а також результати дослідних робіт безпосередньо в свердловині: дані геофізичних досліджень, пробних і дослідних відкачок, режимних гідрогеологічних спостережень, та параметри, що характеризують якість підземних вод. Кластерний аналіз відбувається в просторі ознак свердловини  $x_1, \dots, x_{84}$ , приклади яких наведено в табл. 1.

Таблиця 1  
Параметри гідрогеологічного дослідження

Позначення змінної	Назва параметра	Область значень	№ етапу дослідження	Терм-множини
$x_1$	Віддаленість від населених пунктів, км	0–50	1	{Н – низька, С – середня, В – висока}
$x_2$	Віддаленість від шосейних доріг загальнодержавного значення, км	0–50	1	{Н – низька, С – середня, В – висока}
...				
$x_{84}$	Гідрогеологічні умови за ступенем вивченості	0–10	3	{А – категорії А, В – категорії В, С1 – категорії С1, С2 – категорії С2, Р – категорії Р}

Сформовано навчальний набір даних на основі архівних даних досліджень свердловин родовищ підземних вод, розташованих на території Правобережної Геологічної Експедиції. Вхідному вектору, що містить усі параметри свердловини  $x_1$ – $x_{84}$ , ставиться у відповідність висновок експерта-гідрогеолога про її придатність до видобутку питної води терміном на найближчі 5 років. Навчальний набір даних складається з 20 зразків, приклади зразків наведено в табл. 2.

Таблиця 2  
Дані досліджень свердловин (навчальна вибірка)

Змінна	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$x_1$	1,5	3	12	22	18	15	15	26	37	35	27	4	48	50	50	30	25	22	39	31
$x_2$	43	12	12,5	25	4	11	32	2	5	2	10	3	3	0,8	4,5	1	40	14	1,5	3
...																				
$x_{84}$	5	4	3	2	5	4	4	9	9	7	9	9	8	8	10	10	10	10	10	10

Розіб'ємо навчальний набір даних на кластери за методом [11]. Число кластерів вважаємо наперед заданим,  $c=3$ .

Визначено ступені належності до кожного з кластерів для зразків досліджень свердловин із тестової вибірки. Тестова вибірка складається з 10 зразків, що не входять до навчальної. Результати обчислень ступенів належності зразків тестового набору даних до трьох кластерів подано в табл. 3. Аналіз розташування центрів та складу кластерів в розрізі поняття «перспективність свердловини» дозволяє поставити у відповідність кластерам значення перспективності: кластер 1 – висока, кластер 2 – недостатня, кластер 3 – достатня.

Таблиця 3

Ступені належності зразків тестової вибірки

№	Кластер 1		Кластер 2		Кластер 3		Результат/ширина інтервалу	Оцінка експерта
	Ліва границя	Права границя	Ліва границя	Права границя	Ліва границя	Права границя		
1	0,165	0,178	0,618	0,918	0,249	0,491	Недостатня/0,3	Недостатня
...								
4	0,722	0,99	0,029	0,326	0,053	0,352	Висока/0,27	Недостатня
...								

Тут та далі кольором позначено помилковий вихід системи, тобто такий, що розходиться з висновком експерта для свердловини.

## 5. 2. Модель на основі нечітких множин типу 1

Нечітка логічна системи типу 1 будується як проміжний етап для побудови НЛС типу 2. Вхідний вектор складається зі значень параметрів  $x_1, \dots, x_{84}$ . Базу правил сформовано на основі експериментальних даних із попередніх досліджень артезіанських свердловин (табл. 2). Дані було опрацьовано, поставивши у відповідність кожному значенню з табл. 2 терм-множину нечіткої змінної з табл. 1. Фрагмент отриманої формалізованої бази знань подано в табл. 4.

Таблиця 4

Фрагмент нечіткої бази знань

	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	R <sup>5</sup>	R <sup>6</sup>	R <sup>7</sup>	R <sup>8</sup>	R <sup>9</sup>	R <sup>10</sup>	R <sup>11</sup>	R <sup>12</sup>	R <sup>13</sup>	R <sup>14</sup>	R <sup>15</sup>	R <sup>16</sup>	R <sup>17</sup>	R <sup>18</sup>	R <sup>19</sup>	R <sup>20</sup>
x <sub>1</sub>	В	В	В	С	С	Н	С	С	Н	В	С	В	В	С	С	С	Н	Н	Н	В
x <sub>2</sub>	В	В	С	С	Н	Н	В	В	С	С	С	Н	С	С	С	С	С	Н	Н	Н
...																				
x <sub>84</sub>	В	С1	С1	С2	В	С1	С1	А	А	В	А	А	В	В	А	А	А	А	А	А
y	В	В	В	В	В	В	Д	Д	Д	Д	Д	Д	Д	Д	Н	Н	Н	Н	Н	Н

На повному векторі вхідних даних результат роботи системи співпадає з висновком експерта-гідрогеолога в 20 випадках із 20 для навчальної вибірки та у 8 випадках з 10 для тестової.

Змоделюємо роботу системи типу 1 у випадку недовизначених вхідних даних. Для цього в табл. 1 визначено номер етапу гідрогеологічного дослідження, на якому значення відповідного параметру стає відомим. Поділ на етапи умовний та існує лише для демонстрації результатів; в загальному випадку ставиться вимога роботи системи в довільний момент часу між початком та повним завершенням геологорозвідувальних робіт. Значення параметрів, які стають доступними на більш пізньому етапі, на поточному не враховуються. З антецедент правил бази знань також виключаються стовпці для невідомих змінних. 3-й етап еквівалентний повній визначеності всіх параметрів.

В табл. 5 наведено частину результатів роботи системи на трьох етапах. В цілому помилковий результат було отримано в 6 випадках з 10 на першому етапі, в двох випадках на другому і третьому етапах.

Таблиця 5

Результати роботи НЛС типу 1 на недовизначених вхідних даних

№	Вихід системи			Оцінка експерта
	1-й етап (18 ознак)	2-й етап (35 ознак)	3-й етап (84 ознаки)	
1	4,04	3,52	4,12	недостатня
...				
4	5	5	3,36	недостатня
...				

### 5. 3. Модель на основі нечітких множин типу 2 та отримання агрегованого результату

Експеримент було повторено для моделі на основі нечітких множин типу 2, часткові результати показано в табл. 6.

Таблиця 6

Результати роботи НЛС типу 2 на недовизначених вхідних даних

№	1-й етап (18 ознак)		2-й етап (35 ознак)		3-й етап (84 ознаки)		Оцінка експерта
	Вихід системи	Ширина інтервалу	Вихід системи	Ширина інтервалу	Вихід системи	Ширина інтервалу	
1	[1,39; 4,51]	3,12	[0, 24; 6,27]	6,03	[3,49; 8,97]	5,48	недостатня
...							
4	[1,18; 5,22]	4,04	[1,69; 3,76]	2,07	[1,86; 10]	8,14	недостатня
...							

Результат роботи моделі для одного вхідного вектора тестового набору після пониження простору ознак показано в табл. 7.

Людиною-експертом 1 виділено 39 інформативних ознак, експертом 2 – 41 ознаку. Автоматизованим методом виділення ознак [11] було отримано 32 інфор-

мативні ознаки. Окрім інтервального виходу НЛС, результати також містять виходи двох альтернативних моделей. Альтернативна модель 1 – модель прийняття рішення на основі інтервального кластерного аналізу. Альтернативна модель 2 – рішення, прийняте виходячи з гранично допустимих значень параметрів, контрольованих згідно з законодавством України.

Таблиця 7

Результати роботи агрегованої моделі на недовизначених вхідних даних

№ етапу ГРР	Експерт 1		Експерт 2		Автоматизований метод		Повний вектор				Агрегований вихід
	Визн. ознак/всього	Вихід моделі	Визн. ознак/всього	Вихід моделі	Визн. ознак/всього	Вихід моделі	Визн. ознак/всього	Вихід моделі	Вихід альтерн. моделі 1	Вихід альтерн. моделі 2	
1	9/39	[1,39; 1,51]	8/41	[0,43; 0,45]	9/32	[1,37; 1,42]	18/84	[1,39; 4,51]	х	х	[1,39; 1,51]
2	16/39	[0,24; 0,27]	19/41	[0,01; 0,01]	16/32	[0,09; 0,03]	35/84	[0,24; 6,27]	х	х	[0,24; 0,27]
3	39/39	[0,09; 3,97]	41/41	[0,43; 0,69]	32/32	[0,23; 0,33]	84/84	[3,49; 8,7]	Н:[0,62; 0,92]	недостатня	[3,49; 3,97]

## 6. Обговорення результатів роботи моделі оцінювання стану артезіанської свердловини

Дослідження моделі на основі інтервальної кластеризації зокрема показують розходження рішення, прийнятого системою, з експертним висновком у прикладі 4 (табл. 3). Зразок 4 за всіма показниками крім одного (концентрація радону, 219 Бк/дм<sup>3</sup>) близький до кластера 1. Оскільки досліджуваний метод не має можливостей для врахування інших факторів, окрім Евклідової відстані між точками в просторі ознак, зразок 4 віднесено до кластера 1 (Висока), хоча насправді вода з такими характеристиками непридатна до вживання. Тому система підтримки прийняття рішень у даній задачі вимагає коригування за допомогою експертних знань, чого математичний апарат кластерного аналізу як технологія навчання без учителя забезпечити не може. В решті випадків результат роботи системи узгоджується з рішенням експерта для відповідного зразка; ширину інтервалу можна вважати мірою невизначеності, спричиненої браком експертних знань. Вона досить суттєва, як і слід очікувати від такого складного об'єкта дослідження як гідрогеологічна система.

Що стосується результатів, отриманих НЛС типу 1, то значна частка помилкових розв'язків на початкових етапах дослідження, коли вхідний вектор містить велику кількість пропусків, ще раз підтверджує непридатність механізму побудови нечіткого логічного висновку типу 1 для задач, які допускають наявність пропусків у вхідних даних. Щодо помилок на 3-му етапі, коли вхідні дані повністю визначено, то інструментарій нечітких множин типу 1 не дає змоги визначити їх джерело.

Модель на основі нечітких множин типу 2 дозволяє кількісно оцінити невизначеність, пов'язану з отриманими результатами. На завершальному етапі дослідження, коли значення всіх параметрів відомі, зона невизначеності в багатьох випадках заповнює всю або більшу частину області визначення вихідного параметру. На попередніх же етапах, коли кількість доступних вхідних ознак менша, інтервал невизначеності має меншу ширину, що не має сенсу з точки зору теорії інформації [38]. Цей факт дає змогу зробити висновок, що висока розмірність простору вхідних ознак ускладнює роботу системи.

Невизначеність, притаманна результатам з табл. 6, була також присутня в експериментах із моделлю на основі множин типу 1, але вихід у вигляді єдиного значення не давав можливості пояснити розходження між отриманим та очікуваним результатами. Зважаючи на результати роботи моделі на основі інтервальних нечітких множин, можна зробити висновок, що помилки, отримані при тестуванні моделі типу 1, так само пояснюються надлишковістю простору вхідних ознак.

Реалізація єдиної агрегованої моделі прийняття рішення про стан природної системи має такі переваги порівняно з іншими існуючими підходами:

- узагальнення та повторне використання обчислювальних методів, що використовуються для побудови виходу окремих моделей;
- скорочення проміжку часу, необхідного для прийняття практичних рішень у сфері природокористування, раціонального освоєння та охорони природних ресурсів;
- забезпечення своєчасного прийняття рішень у динаміці природного процесу, та достатньої достовірності прийнятого рішення.

Результати дослідження потенційно можуть бути застосовані в довготривалих програмах вивчення природних процесів задля скорочення терміну прийняття рішення та заощадження необхідних для цього ресурсів. Зокрема в даній роботі показано застосування отриманих результатів на проміжних етапах геологічної розвідки для наближеної оцінки перспективності видобутку підземних вод.

Проведена робота є продовженням досліджень [4, 11, 15, 37] у галузі нечіткого кластерного аналізу та нечіткого логічного висновку, і являє собою зведення результатів згаданих досліджень у завершену систему, готову до застосування в прикладних задачах.

## **7. Висновки**

1. Запропоновано систему підтримки прийняття рішень на основі математичного апарату інтервальних нечітких множин другого типу. На відміну від систем на основі нечітких множин типу 1, результатом роботи яких є ступінь належності в формі єдиного числа, нечіткі множини типу 2 дозволяють отримати на виході інтервал можливих значень вихідної лінгвістичної змінної. Цей інтервал виникає як наслідок невизначеностей, присутніх у вхідних даних, а також невизначеностей, пов'язаних із характером подання експертних знань. Ці особливості дають змогу системі на основі інтервальних нечітких множин працювати в умовах наявності пропусків у вхідних даних, коли робота нечітких логічних систем першого типу неможлива.

2. Побудовано множину моделей, які відрізняються між собою набором входних ознак, виділених експертами з предметної галузі як обов'язкових для врахування. У ході роботи з простором входних ознак кожен експерт відкидає частину їх як надлишкові, нерелевантні або такі, що вносять шум. В результаті виникає підмножина універсальної множини ознак, які утворюють входи інтервальної нечіткої логічної системи та антецеденти правил побудови нечіткого логічного висновку. Таким чином, кожен набір ознак, виділених експертом, породжує окрему модель з інтервальним виходом, чим досягається врахування досвіду експерта в процесі прийняття рішення поряд із інформацією, акумульованою в експериментальній вибірці. Множинність моделей дозволяє також розширити поняття експерта та використовувати наряду з «живими» експертами автоматичні або автоматизовані процедури виділення інформативних ознак.

3. Запропоновано правило побудови агрегованого виходу системи, що дозволяє звести в єдиний інтервал результати роботи множини моделей. Агрегований критерій враховує результат роботи системи на повному входному векторі, а також результати роботи всіх моделей, синтезованих шляхом обмеження простору входних ознак експертами. Таким чином, він являє собою узагальнену інтервальну оцінку стану системи на основі даних, доступних на даний момент, та дає змогу скласти уявлення про невизначеність, пов'язану з прийнятим рішенням. Проміжні результати, тобто виходи окремих підмоделей, мають цінність з позиції дослідження простору входних ознак на предмет залежності прийнятого рішення від факту врахування або неврахування того чи іншого параметра.

4. Показано можливість інтеграції сторонніх моделей на основі інших методів та технологій прийняття рішень. Запропоновано альтернативну модель на основі модифікованого методу кластерного аналізу РСМ з інтервальними ступенями належності. Подання ступенів належності в інтервальній формі дозволяє враховувати та моделювати невизначеності, пов'язані з браком експертних знань. Останнє має особливо важливе значення в контексті кластерного аналізу як технології навчання без учителя.

## Література

1. Бардачев, Ю. Н. Методологическая предпочтительность интервальных экспертных оценок при принятии решений в условиях неопределенности [Текст] / Ю. Н. Бардачев, В. В. Крючковский, Т. В. Маломуж // Вісник Харківського національного університету. – 2010. – № 890. – С. 18–28.
2. Панкратова, Н. Д. Системный анализ и оценивание динамики экологических процессов [Текст] / Н. Д. Панкратова, В. В. Заводник // Системні дослідження та інформаційні технології. – 2004. – № 2. – С. 47–59.
3. Боровский, Б. В. Оценка запасов подземных вод [Текст] / Б. В. Боровский, Н. И. Дробноход, Л. С. Язвин; 2-е изд., перераб. и доп. – К. : Выща шк. Головное изд-во, 1989. – 407 с.

4. Кондратенко, Н. Р. Оцінка перспективності артезіанської свердловини з використанням експертних знань [Текст]: матер. ст. П'ятої Міжн. наук.-прак. конф. / Н. Р. Кондратенко, О. О. Снігур // Інформаційні технології та комп'ютерна інженерія, 2015. – С. 219–221.
5. Минаев, Ю. Н. Методы и алгоритмы решения задач идентификации и прогнозирования в условиях неопределенности в нейросетевом логическом базисе [Текст] / Ю. Н. Минаев. – М.: Горячая Линия-Телеком, 2003. – 205 с.
6. Cai, W. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation [Text] / W. Cai, S. Chen, D. Zhang // Pattern Recognition. – 2007. – Vol. 40, Issue 3. – P. 825–838. doi: 10.1016/j.patcog.2006.07.011
7. De Andres, J. Bankruptcy forecasting: a hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (MARS) [Text] / J. De Andres, P. Lorca, F. J. D. C. Juez, F. Sanchez-Lasheras // Expert Systems with Applications. – 2011. – Vol. 38, Issue 3. – P. 1866–1875. doi: 10.1016/j.eswa.2010.07.117
8. Wang, J. A modified FCM algorithm for MRI brain image segmentation using both local and non-local spatial constraints [Text] / J. Wang, J. Kong, Y. Lu, M. Qi, B. Zhang // Computerized Medical Imaging and Graphics. – 2008. – Vol. 32, Issue 8. – P. 685–698. doi: 10.1016/j.compmedimag.2008.08.004
9. Martyniuk, T. B. Applications of discriminant analysis methods in medical diagnostics [Text] / T. B. Martyniuk, L. M. Kupershtein, A. V. Medvid, A. V. Kozhemiako, W. Wojcik, O. Yuchshenko // Optical Fibers and Their Applications 2012. – 2013. – Vol. 8698. – P. 4. doi: 10.1117/12.2019733
10. Oliveira, J. V. Advances in Fuzzy Clustering and Its Applications [Text] / J. V. Oliveira, W. Pedrycz. – John Wiley & Sons Ltd., 2007. – 435 p.
11. Кондратенко, Н. Р. Інтервальна нечітка кластеризація на основі альтернативних критеріїв якості [Текст] / Н. Р. Кондратенко, О. О. Снігур // Наукові вісті НТУУ «КПІ». – 2012. – № 2. – С. 59–66.
12. Martyniuk, T. B. Formalization of the Object Classification Algorithm [Text] / T. B. Martyniuk, A. V. Kozhemiako, L. M. Kupershtein // Cybernetics and Systems Analysis. – 2015. – Vol. 51, Issue 5. – P. 751–756. doi: 10.1007/s10559-015-9767-0
13. Кондратенко, Н. Р. Діагностика гіпотиреозу на основі нечіткої логіки з використанням інтервальних функцій належності [Текст] / Н. Р. Кондратенко, Н. Б. Зелінська, С. М. Куземко // Наукові вісті національного технічного університету України «Київський політехнічний інститут». – 2003. – № 4. – С. 39–46.
14. Zeng, J. Type-2 Fuzzy sets for pattern classification: A review [Text] / J. Zeng, Z. Q. Liu // 2007 IEEE Symposium on Foundations of Computational Intelligence, 2007. – P. 193–200. doi: 10.1109/foci.2007.372168
15. Кондратенко, Н. Р. Нечітке моделювання в задачі оцінки якості зображень лазерних плям [Текст] / Н. Р. Кондратенко, О. О. Снігур // Системні дослідження та інформаційні технології. – 2014. – № 3. – С. 7–16.



16. Liang, Q. MPEG MBR Video traffic modeling and classification using fuzzy technique [Text] / Q. Liang, J. M. Mendel // IEEE Transactions on fuzzy systems. – 2001. – Vol. 9, Issue 1. – P. 183–193. doi: 10.1109/91.917124
17. Зайченко, Ю. П. Анализ инвестиционного портфеля на основе прогнозирования курсов акций [Текст] / Ю. П. Зайченко, М. Есфандиярфард, А. И. Заика // Вісн. Нац. техн. ун-ту України «КПІ». – 2007. – № 47. – С. 168–179.
18. Liang, Q. Interval type-2 fuzzy logic systems: Theory and design [Text] / Q. Liang, J. M. Mendel // IEEE Transactions on Fuzzy Systems. – 2000. – Vol. 8, Issue 5. – P. 535–550. doi: 10.1109/91.873577
19. Mendel, J. M. Uncertainty, fuzzy logic, and signal processing [Text] / J. M. Mendel // Signal Processing. – 2000. – Vol. 80. – P. 913–933. doi: 10.1016/s0165-1684(00)00011-6
20. Ocampo-Duque, W. Assessing water quality in rivers with fuzzy inference systems: A case study [Text] / W. Ocampo-Duque, N. Ferré-Huguet, J. L. Domingo, M. Schuhmacher // Environment International. – 2006 – Vol. 32, Issue 6. – P. 733–742. doi: 10.1016/j.envint.2006.03.009
21. Gharibi, H. A novel approach in water quality assessment based on fuzzy logic [Text] / H. Gharibi, A. H. Mahvi, R. Nabizadeh, H. Arabalibeik, M. Yunesian, M. H. Sowlat // Journal of Environmental Management. – 2012. – Vol. 112. – P. 87–95. doi: 10.1016/j.jenvman.2012.07.007
22. Yan, H. Adaptive neuro fuzzy inference system for classification of water quality status [Text] / H. Yan, Z. Zou, H. Wang // Journal of Environmental Sciences. – 2010. – Vol. 22, Issue 12. – P. 1891–1896. doi: 10.1016/s1001-0742(09)60335-1
23. Zou, Z. Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment [Text] / Z. Zou, Y. Yun, J. Sun // Journal of Environmental Sciences. – 2006. – Vol. 18, Issue 5. – P. 1020–1023. doi: 10.1016/s1001-0742(06)60032-6
24. Iliev, B. A fuzzy technique for food- and water quality assessment with an electronic tongue [Text] / B. Iliev, M. Lindquist, L. Robertsson, P. Wide // Fuzzy Sets and Systems. – 2006. – Vol. 157, Issue 9. – P. 1155–1168. doi: 10.1016/j.fss.2005.12.014
25. Simeonov, V. Assessment of the surface water quality in Northern Greece [Text] / V. Simeonov, J. A. Stratis, C. Samara, G. Zachariadis, D. Voutsas, A. Anthemidis et. al. // Water Research. – 2003. – Vol. 37, Issue 17. – P. 4119–4124. doi: 10.1016/s0043-1354(03)00398-1
26. Dahiya, S. Analysis of groundwater quality using fuzzy synthetic evaluation [Text] / S. Dahiya, B. Singh, S. Gaur, V. K. Garg, H. S. Kushwaha // Journal of Hazardous Materials. – 2007. – Vol. 147, Issue 3. – P. 938–946. doi: 10.1016/j.jhazmat.2007.01.119
27. Singh, B. Use of fuzzy synthetic evaluation for assessment of groundwater quality for drinking usage: a case study of Southern Haryana, India [Text] / B. Singh, S.

Dahiya, S. Jain, V. K. Garg, H. S. Kushwaha // Environmental Geology. – 2008. – Vol. 54, Issue 2. – P. 249–255. doi: 10.1007/s00254-007-0812-9

28. Dou, C. Transient Ground-Water Flow Simulation Using a Fuzzy Set Approach [Text] / C. Dou, W. Woldt, M. Dahab, I. Bogardi // Groundwater. – 2005. – Vol. 35, Issue 2. – P. 205–215. doi: 10.1111/j.1745-6584.1997.tb00076.x

29. Li, J. An integrated fuzzy-stochastic modeling approach for risk assessment of groundwater contamination [Text] / J. Li, G. H. Huang, G. Zeng, I. Maqsood, Y. Huang // Journal of Environmental Management. – 2007. – Vol. 82, Issue 2. – P. 173–188. doi: 10.1016/j.jenvman.2005.12.018

30. Nobre, R. C. M. Groundwater vulnerability and risk mapping using GIS, modeling and a fuzzy logic tool [Text] / R. C. M. Nobre, O. C. Rotunno Filho, W. J. Mansur, M. M. M. Nobre, C. A. N. Cosenza // Journal of Contaminant Hydrology. – 2007. – Vol. 97, Issue 3-4. – P. 277–292. doi: 10.1016/j.jconhyd.2007.07.008

31. Dixon, B. Applicability of neuro-fuzzy techniques in predicting groundwater vulnerability: a GIS-based sensitivity analysis [Text] / B. Dixon // Journal of Hydrology. – 2005. – Vol. 309, Issue 1-4. – P. 17–38. doi: 10.1016/j.jhydrol.2004.11.010

32. Dixon, B. Groundwater vulnerability mapping: A GIS and fuzzy rule based integrated tool [Text] / B. Dixon // Applied Geography. – 2005. – Vol. 25, Issue 4. – P. 327–347. doi: 10.1016/j.apgeog.2005.07.002

33. Кондратенко, Н. Р. Підвищення адекватності нечітких моделей за рахунок використання нечітких множин типу 2 [Текст] / Н. Р. Кондратенко // Наукові вісті національного технічного університету України «Київський політехнічний інститут». – 2014. – № 6. – С. 56–61.

34. Ивахненко, А. Г. Моделирование сложных систем по экспериментальным данным [Текст] / А. Г. Ивахненко. – М. : Радио и связь, 1987. – 120 с.

35. Кондратенко, Н. Р. Особливості застосування функцій належності типу 2 в системах нечіткої логіки [Текст] / Н. Р. Кондратенко // Наукові вісті національного технічного університету України «Київський політехнічний інститут». – 2016. – № 2. – С. 43–50. doi: 10.20535/1810-0546.2016.2.51636

36. Karnik, N. Type-2 Fuzzy Logic Systems [Text] / N. Karnik, J. Mendel, Q. Liang // IEEE Transactions on Fuzzy Systems. – 1999. – Vol. 7, Issue 6. – P. 643–658. doi: 10.1109/91.811231

37. Кондратенко, Н. Р. Еволюційний пошук інформативних ознак із залученням експерта в задачі оцінки якості артезіанської води [Текст] / Н. Р. Кондратенко, О. О. Снігур // Вісник Вінницького політехнічного інституту. – 2015. – № 3. – С. 96–101.

38. Теория информации и ее приложения (сборник переводов) [Текст] / под ред. А. А. Харкевича. – М.: Физматгиз, 1959. – 328 с.